统信软件技术有限公司
UnionTech Software Technology Co., Ltd.

云原生场景下Ceph的使用

杨晓亮

2023年5月

目录
CONTNETS

01 背景介绍

技术演变

统信软件技术有限公司
UnionTech Software Technology Co., Ltd.

阶段1：服务器

- 烟筒式物理设备管理
- 软件与硬件割裂
- 以“设备”为中心

转变1：
资源自动化
KVM

阶段2：云化

- 统一化资源池化管理
- 软件迁移上云
- 以“资源”为中心

转变2：
应用自动化
Docker

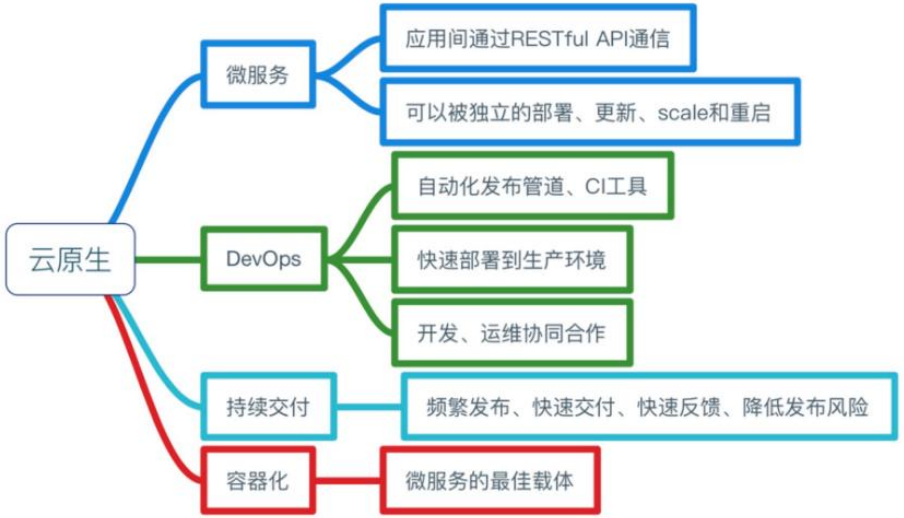
阶段3：云原生化

- 统一云原生基础设施
- 软件云原生架构
- 以“应用”为中心

什么是云原生

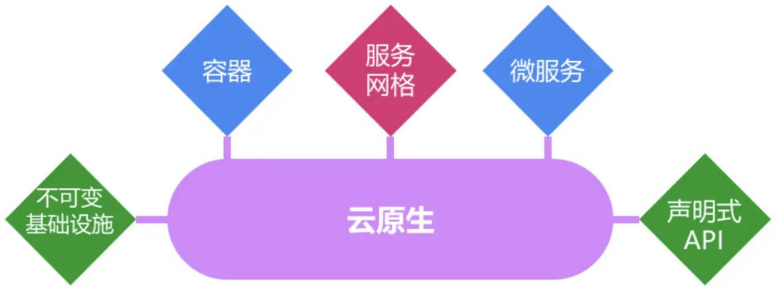
Pivotal 官网对云原生概括为4个要点：

DevOps、持续交付、微服务以及**容器化**。



CNCF定义：

云原生技术有利于各组织在公有云、私有云和混合云等新型动态环境中，构建和运行可弹性扩展的应用。云原生的代表技术包括容器、服务网格、微服务、不可变基础设施和声明式API。



云原生存储的分类

公有云存储	提供一系列云原生存储选项，包括对象存储、文件存储、块存储。
私有云存储	简单可扩展性、高可靠性和便利性特性的商业云存储服务商，提供部署支持和运营与维护 (O&M) 服务。
自建存储	块存储解决方案：Ceph RBD 和存储区域网络 (SAN)；文件存储方案：GlusterFS、NFS 和 CephFS 等。
本地存储	云原生系统中的边缘设备或组件，如数据库、缓存等。

云原生存储发展过程

统信软件技术有限公司
UnionTech Software Technology Co., Ltd.

VP(Volume Plugin)

- 核心代码与K8S主干代码高耦合
- plugin故障导致集群故障
- 开源风险

FlexVolume

- 脚本文件放在host主机上
- 依赖、决兼容问题
- K8S 1.8

CSI （Container Storage Interface)

- 独立性
- 可插拔性
- 高度可定制
- 多存储类型支持
- 安全性
- K8S 1.13

CAS（Container Attached Storage)

- 立即部署
- 智能嵌入K8S

云原生存储方案选择——Ceph

统信软件技术有限公司
UnionTech Software Technology Co., Ltd.

云原生存储特性：

- 高可用
- 可扩展
- 高性能
- 动态部署
- 耐用性

云原生存储挑战：

- 易用性：存储服务部署、运维复杂，云原生化程度低，缺少与主流编排平台整合
- 高性能：大量应用 IO 访问，IOPS 需求高，低时延，性能成为应用运行效率瓶颈
- 高可用：云原生存储已经应用到生产环境，需要高可靠/高可用，不能出现单点故障
- 敏捷性：PV 快速创建、销毁、平滑的扩展/收缩，PV 随 Pod 迁移而快速迁移等

为什么是Ceph

符合云原生存储特性

Ceph生态强于其他开源存储如Glusterfs

克服云原生存储挑战

ceph功能全，支持rbd cephfs ，可以通过网关实现NFS S3接口

02 方案介绍

常见Ceph部署方案对比

统信软件技术有限公司
UnionTech Software Technology Co., Ltd.

ceph-deploy

- ceph14版本之后被弃用
- 非容器化

cephadm

- 容器化部署
- ceph15及更新版本
- systemd+容器
- 更强大的dashboard集成

ceph-ansible

- 容器化部署
- 未适配Orchestrator api
- dashboard部分集成

rook-ceph

- 容器化部署
- 已适配Orchestraort api
- 更强大的dashboard集成
- 对接k8s & ceph首选方法
- ceph14及更新版本

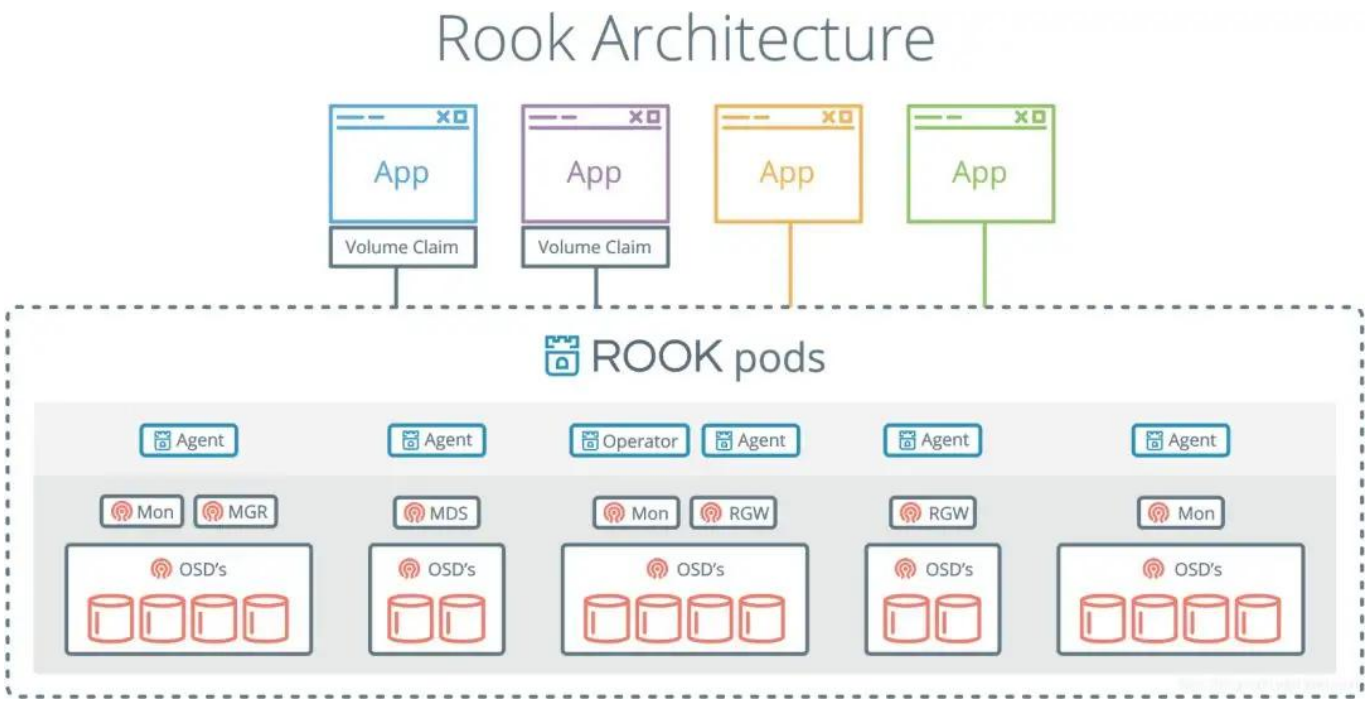
云原生环境Ceph部署方案

Rook-ceph

Rook 是 Kubernetes 的开源云原生存储协调器，为不同的存储解决方案提供平台、框架和支持，以便与云原生环境自然整合。

Rook 将存储软件变成自我管理、自我扩展和自我修复的存储服务。它通过自动部署、启动、配置、配置、扩展、升级、迁移、灾难恢复、监控和资源管理来实现。

Rook 使用底层云原生容器管理、调度和编排平台提供的设施来履行其职责。



Rook-Ceph部署流程

01

Ceph集群部署

rook-ceph-osd 状态为running,视为部署完成。

02

Rook ToolBox安装

Rook 调试和测试的常用工具的容器,Ceph命令行执行端。

03

Dashboard 部署

Ceph集群的状态查看、功能配置。

常见问题及解决

Ceph OSD无法启动

问题：测试环境osd/分区容量过小（小于5G）

解决：替换或扩容至容量大于5G

部署失败后清理集群无法部署OSD

问题：部署失败，清理集群但未格式化磁盘/分区导致部署失败

解决：清理部署集群后，格式化磁盘/分区，重新部署

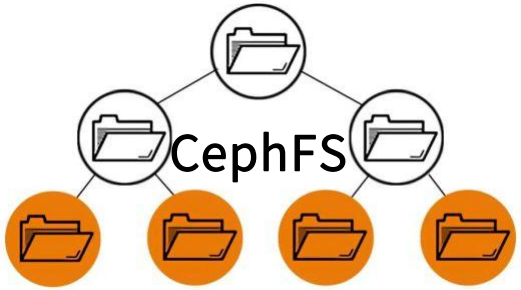
Cephfs pvc一直处于pending

问题：k8s 网络问题

解决：更换k8s 网络组件，或者把ceph集群网络开启host

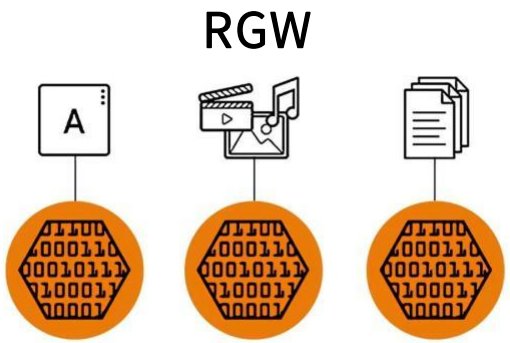
03 使用介绍

Rook-Ceph 存储分类



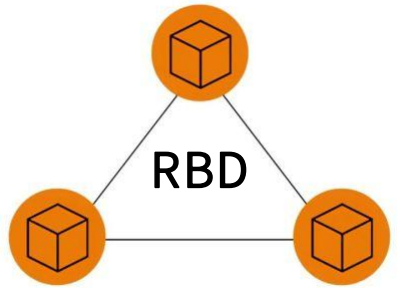
文件存储

可以理解正常文件读写和存储，对应传统存储NAS架构。



对象存储

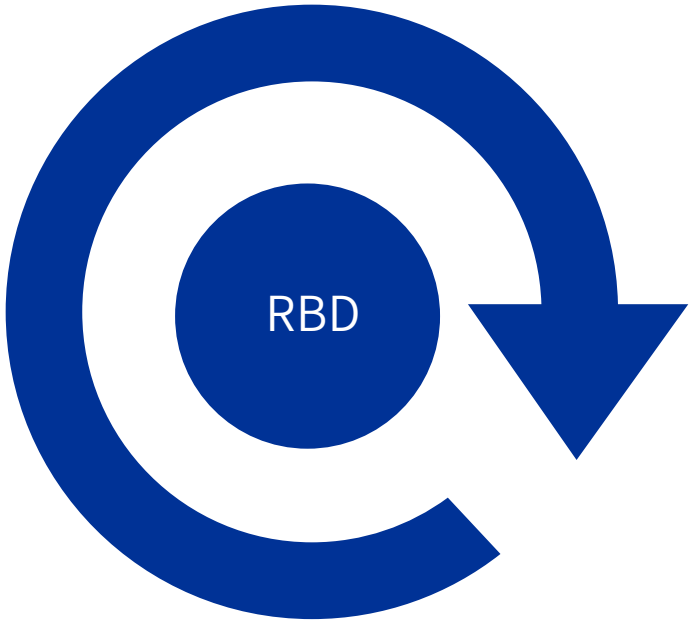
为应用提供 RESTful 类型的对象存储接口。支持S3 Swift接口。



块存储

块存储可以看作成裸盘，不能直接被使用，当挂载到主机后需要指定文件系统。

RBD



1 创建CephBlockPool

创建块存储yaml文件，指定类型为CephBlockPool。

2 创建StorageClass

创建块存储yaml文件，指定类型为StorageClass。

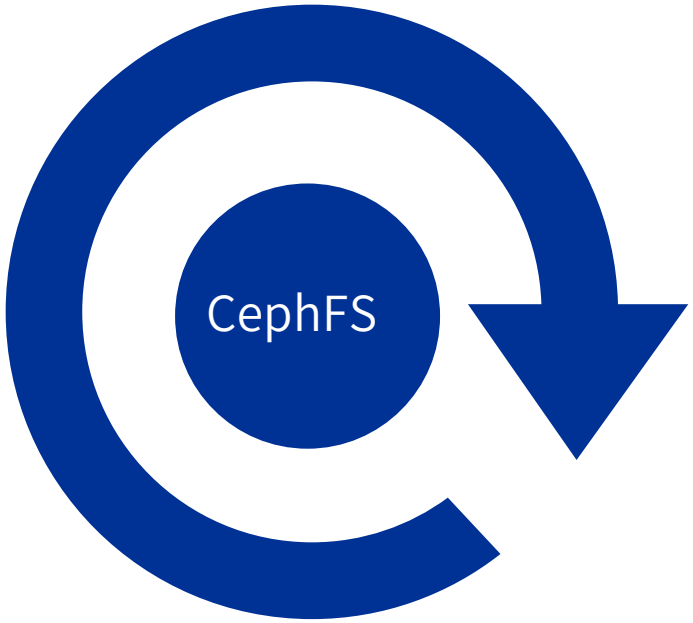
3 Dashboard查看

登录Dashboard查看Pools信息，确认已创建pool信息存在

4 测试验证

创建PVC,创建Pod,挂载申请资源至指定目录，上传文件，其他容器挂载，资源可访问。

CephFS



1

创建 CephFileSystem

创建文件存储yaml文件，指定类型为CephFilesystem。

2

Dashboard查看

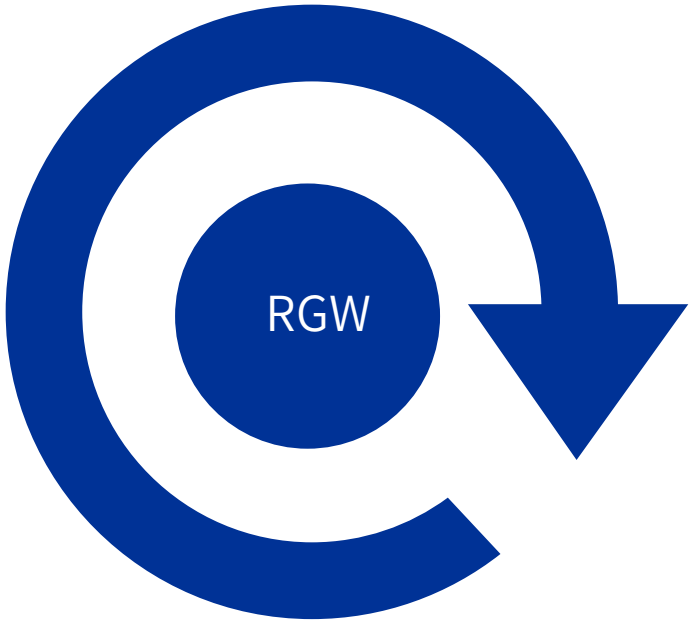
登录Dashboard查看Filesystems信息，确认已创建MDS信息存在。

3

测试验证

创建实例，挂载指定目录，上传文件，其他容器挂载，进入rook-toolbox或Dashboard确认信息。

RGW



1 创建CephObjectStore

创建对象存储yaml文件，指定类型为CephObjectStore。

2 创建User

创建User账户，来生成 AccessKey和SecretKey，为该用户访问 S3 存储使用。

3 集群访问

- 集群内访问：创建bucket--put object--get object。
- 集群外访问：部署一个新的 Service 使用 NodePort 暴露方式。

4 配置 Ceph Object Gateway Management Frontend

直接查看Dashboard Object Gateway,无法查看，提示参考 Ceph Documents 文档，按照参考文档操作，显示信息。

04 案例分享

统信容器云管理平台存储方案应用

统信软件技术有限公司
UnionTech Software Technology Co., Ltd.

使用Ceph分布式存储，提升数据安全性

项目概况

- 以CRI-O、Kubernetes、OKD为基础，以应用为中心的企业级容器云PaaS平台。
- 使用Ceph做分布式存储后端
- 提供自动伸缩、配置管理、资源管理、自动运维等功能
- 实现对容器化应用的全生命周期管理

项目分析

- 对比：分别使用本地存储及分布式ceph后端存储，模拟用户使用场景，对比磁盘损坏，工作节点宕机及损坏，查看并对比客户端访问情况。
- 分析：鉴于本地存储非冗余机制，虽读写效率高于ceph存储，但无法确保数据安全。
- 提升：对比本地存储，增加了冗余机制，可以在磁盘损坏、节点宕机损坏情况下仍确保数据安全、完整。

项目过程

- 基础K8s集群部署
- 以应用为中心的云平台部署
- ceph分部署存储部署及对接

项目成果

- ceph分布式存储的使用实现了存储高可用。
- 充分利用了ceph分布式存储可扩展特性，实现了ceph分布式存储横向、纵向扩展
- 实现了ceph版本的平滑升级

因理想而出生
为责任而成长